VI International Conference on Computational Methods in Marine Engineering
MARINE 2015
F. Salvatore, R. Broglia and R. Muscari (Eds)

IMPROVING PREDICTIVE QUALITY OF KRIGING METAMODEL BY VARIOGRAM ADAPTATION

Daniele Peri

CNR-IAC - Istituto per le Applicazioni del Calcolo "Mauro Picone"
Via dei Taurini, 19 - 00184 Roma, Italy
e-mail: d.peri@iac.cnr.it, web page: http://www.iac.rm.cnr.it/ peri/

Key words: Interpolation/approximation methods, Metamodels, Kriging Interpolation.

Abstract. Application of interpolation/approximation techniques (metamodels, for brevity) is commonly adopted in numerical optimization, typically to reduce the overall execution time of the optimization process. A limited number of trial solution are computed, covering the design variable space: those trial points are then used for the determination of an estimate of the objective function in any desired location of the design space. The behaviour of the prediction of the objective function in between two trial points depends on the structure of the adopted metamodel, and there is no possibility, in principle, to determine *a priori* if one method is preferable to another. Nevertheless, some metamodels require the adjustment of a set of tuning parameters, and this operation is critical for the prevision qualities of the metamodel. In this paper, some base parameters of the kernel of the kriging metamodel are tuned in order to improve the overall quality of the prediction.

1 Introduction

The substitution of the true objective function with an interpolation/approximation method (say, a metamodel) is largely adopted in optimization: this technique is particularly efficient (sometime indispensable) when the objective function is computationally expensive, i.e. it is obtained as a result of complex numerical simulations, and the time for completing the design activities is not compatible with the use of the true objective function in every phase of the optimization process. In this case, reliability of the metamodel becomes crucial, and the use of unreliable metamodels may led to the lost of the true optimum of the function.

Among the different types of metamodelling techniques, kriging represent a really attractive methodology: determination of the base parameters of the metamodel is cheap, if compared i.e. with neural networks, and the objective function value is preserved on the points where the true value of the objective function is known (training points). The behaviour of the interpolation in between the training points is determined by the use

of the so-called *variogram*, that is, an algebraic function expressing a spatial correlation between two points. An approximation of the variogram can be obtained by using the training points: unfortunately, rarely the coverage of the design space is dense enough to be statistically significant, so that the resulting variogram is unreliable. As a consequence, the variogram is commonly substituted by an algebraic function, typically from the family of the radial basis functions, fitting the experimental variogram obtained from the training points.

Different assumptions for the variogram led to different behaviors of the metamodel onto the design space, so that it is reasonable to think that an optimal selection of the shape of the variogram, able to minimize the average predictive error over a finite portion of the design space, may exists.

In this paper, some techniques for the adaptation of the variogram to the fitting function are presented and discussed. Depending on the adopted strategy, some indicators potentially expressing the predictive qualities of the kriging metamodel are minimized through dedicated optimization algorithms. Although a complete correlation between the average error and the selected indicators has not been observed, a reduction of the global predictive error has been obtained.

2 Some base elements of Kriging

If a set of sampled data of the objective function is available, say

$$Z = (z(s_1), z(s_2), ...z(s_N))^T$$

where z is the objective function value and s_i is the sampling point, they can be viewed as observations of a Gaussian random process

$$\{Z(s) : s \in D\}; D \in \Re^N$$

If we assume for Z the expression

$$Z(s) = \mu + \delta(s); s \in D$$

where μ is the expected value of the function generating the sampled set Z and δ is a zero-mean stochastic process with known covariance function

$$C(s, u) = cov(Z(s), Z(u)); s, u \in D$$

it can be demonstrated that an approximation of Z can be expressed as

$$Z^*(s) = c'C^{-1}Z + (1 - c'C^{-1}I)\mu$$

where C is the covariance matrix $(N \times N)$ $C(s_i, s_j)$, c' the covariance vector $\{C(s_0, s_1), C(s_0, s_2), ..., C(s_0,$
I the identity matrix.

Since a Gaussian process is hypothesized, mean-squared prediction error is equal to

$$Err(Z(s_0) - Z^*(s_0)) = C(s_0, s_0) - c'C^{-1}c$$

This is the so-called *Simple Kriging*. A variant of simple kriging, widely used in literature (and adopted in the following), is the so-called *Ordinary Kriging*. Ordinary Kriging is similar to Simple Kriging, but a further equation is enforced, so that the sum of the resulting weights iλ_i is equal to the unit value:

$$\begin{aligned} \sum_{j=1}^{N} \lambda_j \sigma_{ij} + \mu &= z(s_i) \quad i = 1, ..., N \\ \sum_{j=1}^{N} \lambda_j &= 1 \end{aligned}$$

The predicted value is simply obtained as the weighted sum of the known values

$$Z^*(s_0)) = \sum_{i=1}^{N} \lambda_i z(s_i)$$

and the associated prediction error is here

$$Err(Z(s_0) - Z^*(s_0)) = C(s_0, s_0) - \mu - \sum_{i=1}^{N} \lambda_i C(s_0, s_i)$$

It is evident how the method is founded on the availability of the covariance function, or a reliable approximation of it. In practice, μ is replaced by the average of the available samples, and the covariance function is typically substituted by the semi-variogram γ defined, in discrete form, as

$$\gamma(h) = \sum_{i=1}^{n(h)} \frac{(z(s_0 + h) - z(s_0))^2}{n(h)}$$

where $z(s_0)$ denotes the data value at a particular location, h is the distance between two locations, and $n(h)$ is the number of pairs of data values a distance of h apart.

Unfortunately, the experimental semi-variogram is, in reality, a point function, so that it changes when s_0 spans the design variable space: as a consequence, the prediction of the objective function Z^* will be not exact on every point, if a single covariance function is adopted, while the value is preserved at the sample points, by construction.

In practical applications, algebraic functions are adopted as covariance function. A common selection is the exponential variogram

$$C(s_1, s_2) = \gamma(h) = c_0 + c_1 \left(1 - exp\left(\frac{h}{r}\right)\right); h = ||s_1, s_2|| \quad (1)$$

Parameters c_0 and c_1 can be tuned on the base of the available training set.

A final annotation about the structure of the algebraic functions typically adopted as semi-variograms. In practice, adopted correlation functions are spherical, since he definition of the variogram is spheric, so that they do not depend on the specific mutual position of the computing points but their distance. As a consequence, the sensitivity of the semi-variogram is the same for each design variable, and there are not preferential directions. This is not what is observed in practice, in particular in the field where kriging has been developed, that is, statistical geology.

3 Variogram adaptation technique

A common technique for the determination of the variogram is represented by fitting an algebraic function onto the experimental values of the variogram, obtained from the available samples $Z(s_i)$, and then adopting the fitting function as variogram [2]. By this approach, we are implicitly accepting the hypothesis that the sample data are statistically significant in the description of the real variogram of the function Z, that is, they are representative of the variogram of the fitting function. As an alternative, another interesting technique is the one generating the so-called *minmax kriging* [7]: here a large class of plausible variograms is adopted, minimizing the maximum possible kriging variance in that class. This approach requires a quite large number of attempts, but the price may be still not comparable with the effort required for the evaluation of the objective function at a point, so that it is, in many practical cases, absolutely feasible. This approach allows the consideration of a variety of families for the variogram, but a more extensive analysis should be performed in order to determine the best parameters for each class of variograms: in fact, a better solution could be lost due to under-sampling of the parameter space.

In this paper, the approach is different than the classical variogram fitting, much more similar to [7]: we assume that a prediction of the global error is possible, instead of considering the fitting properties of the variogram as the only way to judge the correctness of the prediction. Consequently, we define a quality indicator (QI), possibly strictly correlated with the overall prediction qualities obtained by two different metamodels: QI is minimized by using a nonlinear optimization algorithm, acting on the control parameters of the variogram. If the correlation between QI and the overall prediction qualities of the metamodel is good, we are producing an improved metamodel.

As a measure of the quality of the metamodel, the average error (AE) is defined. AE is computed comparing the predicted value with the true value of the objective function on a large number of sample points, different from the ones forming the training set. AE is performed using a regular sampling of the space (101x101). This will be adopted during the optimization procedure in order to check the correlation between QI and AE, but

this will be not driving the optimization process in any way, since this is not feasible in practical applications.

Some preliminary studies have been already presented in [6], where two different choices for QI have been adopted:

1. self-correlation (SC): if N samples are available, N metamodels are obtained by eliminating one point from the training set in turn. The same variogram is adopted for all the N variants. The missing point can be used to compute the prediction error $(Z(s_i) - Z^*(s_i))$. The sum of all the prediction errors obtained by all the N variants of the metamodel, is used as indicator of the predictive qualities of the adopted variogram.

2. training/validation (TV): a subset of the original training set is eliminated, and used as *validation set*. The sum of the prediction errors on the validation set is adopted as indicator.

In order to improve adaptability, N different variograms are used, one for each sample point[1]. Furthermore, scaling factors and rotations around the axis are applied to each design variable, in order to include anisotropic effects. This way, generality of the variogram is increased: as a counterpart, the number of variables is growing. Rotation is introducing some preferential directions for the mutual influence of the sample points: this is a realistic situation in a large variety of fields of application, like in geological statistics, the field in which kriging has been developed.

As a variogram base function, here the investigation is restricted to exponential type only, in which only two free parameters are present (see equation 1). In previous paper [6] the variogram was also modeled by using a spline curve: in order to gain some insight on the general methodology, here the investigation is reduced to a single (simpler) variogram family.

Optimal value of QI is obtained by using two different optimization algorithms in sequence. A fast detection of the basin of attraction of the minimizing function is obtained by applying Particle Swarm Optimization (PSO) algorithm [3]: once the swarm converges such that the search is restricted into a small area, a *pattern search* algorithm [4] is started from the current best point in order to dig into the selected basin of attraction. The use of a sequential approach is suggested by the large number of design variables of the problem: in a simple three dimensional problem, 8 variables for each variogram (sample points) are needed, two for the exponential function parameters, one for each scale factor and one for each rotation. This number is to be multiplied by the number of available sample points, so that it is very easy to face problems with hundreds of parameters to be managed and minimized. Further details can be found in [6].

[1] note that the use of multiple variograms causes the loose of the biunique correspondence between two generic points in space, so that $C(s_i, s_j) \neq C(s_j, s_i)$)

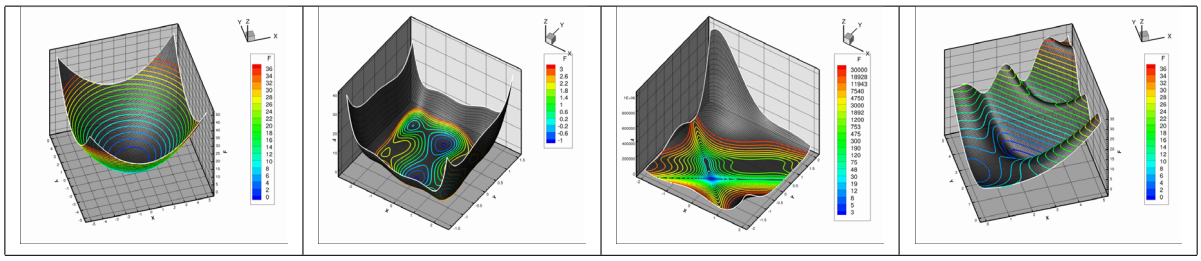

Figure 1: 3-D representation of the four algebraic functions adopted for the test. From left to right, Spheric, Six-Humps Camel Back, Goldstein-Price, Sasena.

4 Numerical results

The analysis started in [6] has been here extended to four different algebraic functions in the two dimensional space:

1. Spheric function: one single global minimum. Equation:

$$y = x^2$$

2. Goldstein-Price Function: one single global minimum, but complex valley. Equation:

$$\begin{aligned} y &= (1 + (x + y + 1)^2(19 - 14x + 3x^2 - 14y + 6xy + 3y^2)) \times \\ &\quad (30 + (2x - 3y)^2(18 - 32x + 12x^2 + 48y - 36xy + 27y^2)) \end{aligned}$$

3. Six-Humps Camel Back function: six minima, two global minima. Equation

$$y = \left(4 - 2.1x^2 + \frac{x^4}{3}\right) x^2 + xy + (-4 + 4y^2)y^2$$

4. Sasena Function: multimodal, one global minimum in a limited set. Equation

$$\begin{aligned} y &= 1.2 + 0.05\left(0.01(y - x)^2 + (1 - x)^2 + 2(2 - y)^2\right) + \\ &\quad 0.05\left(7sin(0.5x)sin(0.7xy)\right) \end{aligned}$$

Peculiarity of these function (but the first) is the difficulty to capture the main features with a few samples.

4.1 Quality index: Self-Correlation and Training/Validation

Application of previously defined QI (both SC and TV) failed to identify a correlation between QI and AE. In other terms, the reduction of the QI is mostly resulting in an increment of the AE, and not *viceversa*. Examples are reported in figure 2 and 3.

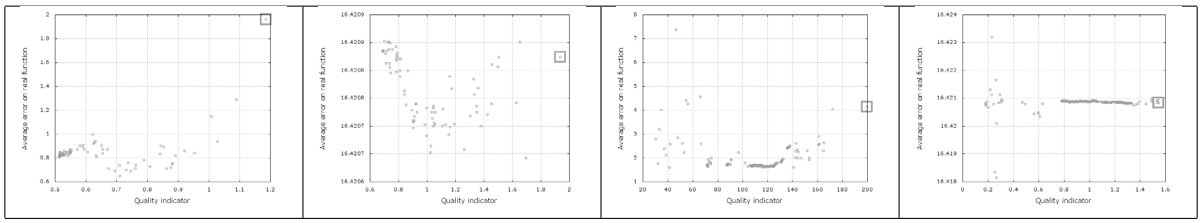

Figure 2: Effectiveness of the SC quality index using a training set of 100 sample points: a reduction of the QI should be reflected on the improvement (reduction) of the AE. From left to right, results obtained using Sferic, Six-Humps Camel Back, Goldstein-Price and Sasena functions. Square dot reports the initial point (without optimization).

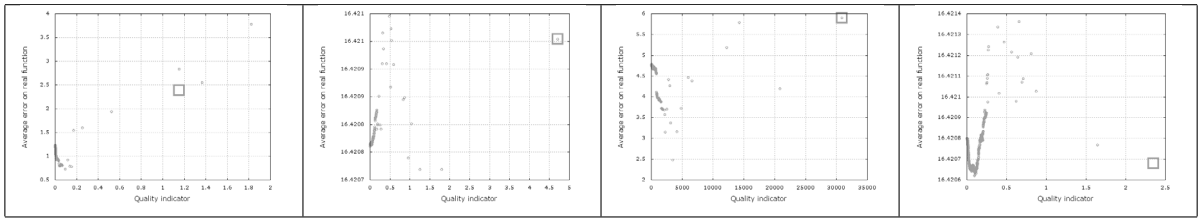

Figure 3: Effectiveness of the TV quality index using a training set of 100 sample points: a reduction of the QI should be reflected on the improvement (reduction) of the AE. From left to right, results obtained using Sferic, Six-Humps Camel Back, Goldstein-Price and Sasena functions. Square dot reports the initial point (without optimization).

In figure 2 the effect of the minimization of the SC indicator for the four test functions is reported. The square spot is representing the starting point: at the end of the optimization of the QI, one would expect a reduction of the AE. Sferic function is the only reporting a decrease of both SC and AE together: this is probably due to the fact that a single basin of attraction is present. On the contrary, for all the other functions, the reduction of SC is causing an increase of AE, or an uncertain value in the best case. The same situation is observed for the TV indicator, as can be observed in figure 3. In the case of the Sferic function, SC appears to be preferable to TV.

Poor results are suggesting the definition of a different QI.

4.2 Quality index: Average Prediction Error

After this first round of tests, a further QI has been adopted. Exploiting another feature of kriging, that is, the capability to provide an estimate of the prediction error, we can produce an average of the prediction error (APE) to be used as QI. This is, in principle, what is already proposed in [7]. The main differences are:

1. the systematic exploration of the space of parameters by numerical optimization algorithms, and

2. the use of a variety of variograms, with space adaptation and anisotropic effects.

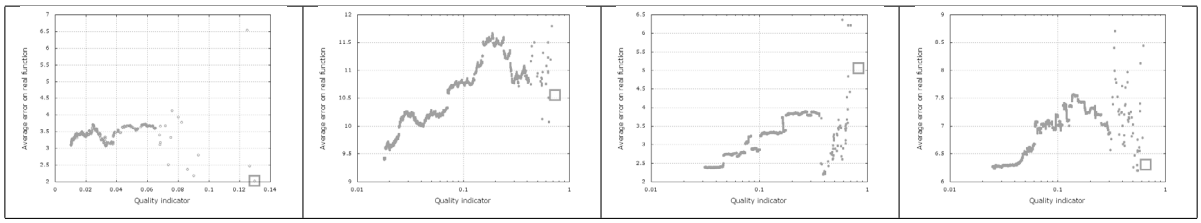

Figure 4: Effectiveness of the APE quality index using a training set of 49 sample points: a reduction of the QI should be reflected on the improvement (reduction) of the AE. From top to bottom, left to right, results obtained using Sferic, Six-Humps Camel Back, Goldstein-Price and Sasena functions. Blue dot reports the initial point (without optimization).

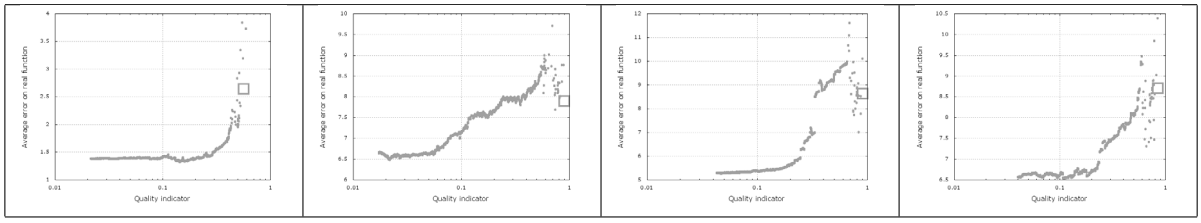

Figure 5: Effectiveness of the APE quality index using a training set of 121 sample points: a reduction of the QI should be reflected on the improvement (reduction) of the AE. From top to bottom, left to right, results obtained using Sferic, Six-Humps Camel Back, Goldstein-Price and Sasena functions. Blue dot reports the initial point (without optimization).

Being the prediction error higher far from the training points, APE is evaluated at the centroids of the hyper-tetrahedron obtained by a Delaunay triangularization of the training set.

APE appears to be much more correlated with AE, and the results reported in figures 4 and 5 are encouraging. In figure 4, the minimization of APE based on a training set of 49 points is reported. The final point presents, without exceptions, a reduced value of AE, although in some cases a further reduction on the QI is not producing a reduction on AE. Small improvements are obtained on the more complex function.

If the training set is further enriched, passing from 49 to 121 sample points, the results appears to be more stable and coherent. In figure 5 we can observe how an improvement in the predictive quality of the metamodel is obtained for every test function, being the improvements well perceptible, also on the more complex Sasena function.

5 Discussion

Kriging methodology is founded on the availability of a covariance function, providing the spatial correlation between two points. This has been largely illustrated in section 1. The variogram is implicitly determining the shape of the function in between two training points: a single variogram is commonly adopted for the entire design space, but it is rather intuitive to hypothesize a change of a the shape of the variogram at different locations, so that a single function is not able to reproduce the real behaviour of the covariance

function.

Since the variogram is, at last, the main responsible for the correct interpolation of the objective function, the general idea inspiring this work was to act directly on the variogram in order to improve the prediction capabilities.

The use of the prediction error as a criterion for the ranking of different variograms, as performed in [7], is equivalent to act on a derived quantity, and not directly on the real difference between prediction and true value. As a consequence, this approach was not considered at the beginning, and the use of a QI aimed at the modeling of the true difference between prediction and objective function was preferred. The general idea was to utilize a subset of the sample points in order correct the variogram, driving the interpolated values onto the true value of the objective function in those regions where training points are not present.

Evidently, the improvement of the fitting qualities of the model in correspondence to the samples from the validation set is also increasing the risk of overshooting in some locations different than the training and validation points, so that the overall predictive qualities of the improved metamodel are deteriorated from the initial guess. On the contrary, the use of the predicted error appears to provide good results in reducing the average error. This is in some sense unexpected, considering that the expression of the prediction error is obtained under some hypothesis that are not respected by the optimal variograms.

A more deep investigation about the effects of the minimization of the prediction error is required, in order to better understand the meaning of the obtained results.

6 Conclusions

A methodology for the determination of some criteria for the improvement of the quantitative prediction qualities of kriging metamodel has been proposed. Preliminary results obtained on a small suite of test function provide encouraging results for the proposed approach, although some aspects are still unclear. Further numerical tests are required in order to obtain a stronger validation, possibly providing some more insights on the interpolation mechanism.

REFERENCES

[1] Chil J.P., Delfiner P., 'Geostatistics - Modeling Spatial Uncertainty'. Wiley Series in Probability and Statistics, Second edition, 2012.

[2] Cressie N., 'The Origins of Kriging'. *Mathematical Geology*, Vol. 22, No. 3, 1990.

[3] Kennedy, J., Eberhart, R., 'Particle Swarm Optimization'. *in Proceedings of IEEE International Conference on Neural Networks*, IV. pp. 19421948, 1995.

[4] Lewis R.M., Torczon V.. 'Pattern search methods for linearly constrained minimization'. *SIAM Journal on Optimization*, Vol. 10, Issue 3, pages 917-941, 2000.

[5] Peri D., 'Self-Learning Metamodels for Optimization'. *Ship Technology Research*, 56(3):94-108., 2009.

[6] Peri D., 'Automatic Tuning of Metamodels for Optimization'. *In COMPIT'13 - 12th International Conference on Computer and IT Applications in the Maritime Industries*, Cortona (AR), Italy, 15-17 Aprile 2013.

[7] Pilz, J., G. Spoeck, M. G. Schimek, 'Taking account of uncertainty in spatial covariance estimation'. *In Geostatistics Wollongong 96*, Vol. 1, E. Y. Baafi and N. A. Schofield, eds. Kluwer, Dordrecht, pp. 302-313, 1997.